# FAQ sull'inutilità dei ritardi nel lotto (e altrove)

di Adam "bad attitude" Atkinson

February 10, 1999

**Abstract**

Questa vorrebbe essere una FAQ sul perchè la tesi ritardista (conviene giocare i numeri che mancano da più tempo nella lotteria) sia incoerente e/o balorda e/o impossibile da giustificare. Questa versione è incasinata: sono stato fuori per due settimane, e ho parecchie cose da aggiungere. Non so bene dove in che ordine metterle. Allora, considerate questa una versione alpha incompleta.

## 1 Introduzione

Abbastanza spesso, qualcuno in `it.fan.dewdney` , `it.scienza.matematica`, `it.hobby.enigma` o `it.scienza` cerca di dire che sia utile conoscere i "ritardi" (o qualche altra informazione sui risultati passati) per scegliere i numeri del lotto.

Come tutte le FAQ, questa FAQ è stato scritta perché io non debba più scrivere queste spiegazioni in area ogni volta che arriva un altro ritardista: potrò limitarmi a dire "Sigh... leggiti la FAQ." Poi, magari sarà utile a chi si trova a discutere con i ritardisti ma non sa bene come combattere i loro cosiddetti "ragionamenti".

Qualche commento su alcune delle mie scelte controverse (almeno secondo commenti ricevuti finora):

Ho scritto il FAQ in LaTeX perché l'`ASCII` e l'`HTML` non sono molto adatti per le formule, e il formato più usato per i documenti matematici/tecnici/scientifici è il `postscript`. Visualizzatori postscript si trovano su `http://www.cs.wisc.edu/ ghost/index.html`, e sono disponibili per molti sistemi operativi diversi. Parlando di sistemi per scrivere documenti matematici, Tom Körner dice "It is, in my belief, a waste of time to learn anything but a TeX-based system", e mi trovo d'accordo. Non ci sono alternative credibili. Il Word non è credibile per vari motivi, fra cui il fatto che non ho Word, e molte altre persone non possono leggere i file Word. Sulla home page di questa FAQ (`http://www.mistral.co.uk/ghira/ritardi.html`) ci saranno, per ora, versioni TeX, `dvi` e `postscript`. Nota che non sono molto esperto in TeX; Elio Fabri ha gentilmente corretto i miei errori più madornali. Sulla home page ci sono anche dei programmi in `perl` e/o `C`.

Mi è stato detto che se uso le formule, qualcuno potrebbe o non capire il documento, o rifiutare di leggerlo. *R*ifiuto di togliere le formule solamente perchè qualche povera stellina sensibile rischia di svenire se vede una $x$. Se non ti piacciono le formule, puoi anche saltarle: nella maggior parte di questo documento, non sono essenziali. Vorrei che questo documento fosse di qualche utilità a persone con background abbastanza vari, ma questo (IMHO) non vuol dire togliere ogni cosa che almeno un potenziale lettore potrebbe non capire. Circa 10 anni fa, mi è stato detto che alcune persone non capiscono le frasi con le virgole in mezzo, perché sono complicate. Andando avanti così, cercando di eliminare ogni possibile fonte di difficoltà, arriviamo a:

> Io volere essere compreso. Io scrivere frasi brevi. Io usare molti esclamativi! Così testo sembrare più divertente! Io usare molte figure colorate!

Poi, non mi piace l'atteggiamento di alcune persone che le formule sono superflue, o un codice segreto usato dai matematici per nascondere quello che fanno, ecc. ecc. I matematici usano le formule per comunicare le loro idee, non per nasconderle. A cosa serve dire "certi fenomeni si verificano entro certo limiti", come fa Ritardista di Turno #2, quando posso dire esattamente ciò che intendo?

Ci sono anche dei programmi brevi all'interno di questo documento, che servono per illustrare qualcosa. (I programmi dovrebbero essere disponibili come file a parte sulla mia pagina web). Se non capisci i programmi, saltali. Stanno lì per chi è in grado di capirli e/o usarli. Quasi sicuramente tu capirai o apprezzerai qualcosa in questo documento che qualcuno altro avrebbe voluto eliminare perché troppo difficile. Molti programmi sono in perl perché è il mio linguaggio preferito, ed è disponibile, gratis, per molti sistemi operativi. Alcuni programmi erano troppo lenti in perl, e li ho rifatti in C, un altro linguaggio disponibile gratis per molti sistemi operativi.

In inglese avrei scritto questa FAQ in modo non-gender-specific. Ma scrivere ogni volta "il/la ritardista" in italiano non mi va. Quindi questo documento è maschilista. Questa FAQ *dovrebbe* essere comprensibile a chi abbia studiato un po' di probabilità. Nei licei qui in UK la probabilità viene studiata, ma forse non fa parte del programma normale in Italia. Comunque, un qualsiasi studente universitario del primo anno in una materia scientifica non dovrebbe avere problemi. E cercherò di essere abbastanza esauriente, in modo che un liceale dovrebbe poter seguire i passaggi. Cercherò di usare distribuzioni discrete anziché continue, e così via. Usare roba "da liceo" non mi sembra così irragionevole.

Pare che ad almeno due ritardisti, non piace vedere i loro nomi sulla mia pagina web, o in questo documento. Allora ho tolto i loro nomi. Gorgia mi sembra un bersaglio leggittimo, però, e il suo nome rimane.

Avrei dovuto scrivere "con probabilità 1" o "quasi sempre" o "in probabilità" o "in mean square" un sacco di volte in questo documento. Forse ne manca qualcuno. Notate che un evento con probabilità 0 non è necessariamente impossibile: la probabilità che un numero reale scelto uniformemente da $[0, 1]$ sia 0,5 è 0, ma non è impossibile. Nello stesso modo, un evento con probabilità 1 non è necessariamente certo. Chi vuole saperne di più dovrebbe seguire un corso sulla teoria della misura.

I ritardisti del tipo più comune (li chiamerò "indipendentisti": quelli che credono che le estrazioni del lotto siano eventi indipendenti, ma di saper scegliere numeri vincenti più spesso degli ingenui che scelgono numeri qualsiasi) solitamente giustificano la loro posizione con una combinazione di:

1. (i) Cose sbagliate, cioè cose che non sono vere.

2. (ii) Cose vere ma irrilevanti, cioè cose che sono vere, ma che non sostengono la tesi del ritardista. Per esempio, in "7 è maggiore di 5, dunque è un numero primo", "7 è maggiore di 5" è vero ma irrilevante: 7 è maggiore di 5, ma questo non implica che sia un numero primo. In realtà, 7 **è** un numero primo, ma non lo puoi dedurre dal fatto che è maggiore di 5. Anche 8 è maggiore di 5, ma 8 non è un numero primo. Un altro esempio di una cosa vera ma irrilevante potrebbe essere "Atkinson è uno stronzo antipatico arrogante strafottente sprezzante": sarà anche vero, ma di poca utilità in una dimostrazione matematica.

3. (iii) Cose così incoerenti (nel senso di *incoherent*, non *inconsistent*) che è difficile dire se sono vere, false o semplicemente balorde. Ne troverete ottimi esempi in *Fads and Fallacies in the Name of Science* di Martin Gardner.

4. (iv) *Namedropping* — non so come si dice in italiano. Comunque, consiste nel nominare personaggi e teoremi altisonanti nella speranza che dia credibilità al resto del discorso. Per esempio: "Bla bla bla. Teorema di Gödel. Bla bla bla. Euler. Bla bla bla. Hardy e Ramanujan." Uno dei miei corrispondenti propone "Cazzeggio altisonante" come traduzione italiana.

Le cose di cui i ritardisti indipendentisti sembrano abusare, o non capire, sono, solitamente, la (meglio, una: ci sono varie versioni) Legge dei Grandi Numeri e la teoria della probabilità condizionata. Alcuni ritardisti citano il Teorema del Limite Centrale, ma sembrano pensare che sia la Legge dei Grandi Numeri.

Parliamo brevemente dell'indipendenza. Se un ritardista dice che delle prove sono indipendenti, ma che lui comunque può prevedere i risultati futuri grazie ad una conoscenza di quelli passati, la cosa è semplice: possiamo tranquillamente dire che il ritardista sta raccontando balle. Non è una questione di "metodo scientifico" o altri termini altisonanti (vedi *namedropping*, è una semplice questione di definizioni. Due o più eventi sono indipendenti se una conoscenza dei risultati di alcuni non ti aiuta affatto a prevedere i risultati degli altri. Questa è la definizione di indipendenza, almeno per i nostri scopi. Beh, la definizione che si trova su un libro sarà piena di formule del tipo $P(A \cap B) = P(A)P(B)$. Va bene, scateniamoci:

Una famiglia $\{A_i : i \in I\}$ si chiama *indipendente* se

$$P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i)$$

per ogni insieme finito $J$ di $I$.

In particolare, l'indipendenza ci permette di dire

$$P(A|B) = P(A \cap B)/P(B) = P(A)P(B)/P(B) = P(A)$$

ed è principalmente questo che faremo.

Dunque, chi dice "Le estrazioni del lotto sono indipendenti, ma io posso usare delle informazioni sui risultati passati per prevedere quelli futuri meglio di una strategia che ignora i risultati passati" si contraddice. (Queste "informazioni" saranno tipicamente i famosi ritardi). Alcuni ritardisti sembrano avere le idee poco chiare sulla definizione di ritardo, o forse cambiano la definizione quando si sentono sotto pressione dagli stronzi elitari antipatici superbi strafottenti arroganti come me; ma non è importante come sono definiti i ritardi: se partiamo con l'idea che le estrazioni del lotto, (o i lanci di uno o più dadi o altre cose simili) sono indipendenti, allora non importa come definisci i ritardi. Non funzioneranno, perché hai già detto che non serve a niente conoscere i risultati passati. Leggiamo insieme ad alta voce: se gli eventi sono indipendenti, vuol dire che eventuali informazioni sui risultati passati non ti aiutano. I ritardi sono informazioni sui risultati passati. Perciò i ritardi non ti aiutano.

Come Elio Fabri ha giustamente notato, per il caso "ritardista indipendentista" dovrei potermi fermare qui: un tale ritardista si contraddice, e non sa cosa sta dicendo, punto e basta. Comunque, questo è un FAQ, e secondo me è utile fornire risposte alle varie obiezioni sciocche dei ritardisti "indipendentisti" che cercano di inventare storielle sul perchè il nero e il bianco sono, in realtà, la stessa cosa, 3 è maggiore di 5, e ci sono 47 nuovi interi fra 0 e 1. (I ritardisti "dipendentisti" li discuto dopo; la loro posizione è difendibile, anche se con molta difficoltà. È solmanente implausibile, inverosimile e non sostenuta dai fatti, invece di essere una contraddizione.)

## 2 I ritardisti "indipendentisti"

Notate che qui non stiamo parlando del lotto, ma di un modello matematico del lotto. Se non ti piace il modello, non usarlo.

### 2.1 Probabilità Condizionata

I ritardisti spesso dicono cose del tipo "È molto improbabile che una moneta esca testa duecento volte di fila. Perciò se esce testa 199 volte di fila vuol dire che la prossima volta uscirà quasi sicuramente croce". (Supponiamo che la moneta sia *fair*). La risposta a questo è "Ma è altrettanto improbabile che una moneta esca 199 volte testa e poi croce. Infatti, ogni possibile successione di 200 risultati è improbabile." Il ritardista sta dicendo cose vere ma irrilevanti. ("La probabilità che ci siano due bombe su un solo aereo deve essere bassissima. Allora, se porto sempre una bomba quando viaggio, sarò ok." dicono i ritardisti.)

La probabilità di avere 200 teste di fila non c'entra. Quello che vogliamo sapere è la probabilità di avere una testa dato che abbiamo appena avuto 199 teste. Visto che le prove sono indipendenti, la probabilità è quella di avere una testa con una moneta vergine. Non è necessario fare calcoli, ma li faccio comunque:

Supponiamo che in ogni tentativo si abbia un successo con probabilità $p$ o un insuccesso con probabilità $q = 1 - p$, e che i lanci siano indipendenti.

Allora, la probabilità di una data successione di risultati di $a$ successi e $b$ insuccessi è $p^a q^b$.

Vogliamo conoscere

Pr(successo al tentativo $n$ | risultati dei tentativi $1, \ldots, n-1$). Usiamo la definizione della probabilità condizionata:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Allora abbiamo

$$\begin{aligned} &\Pr(\text{successo al tentativo } n \mid \text{risultati dei tentativi } 1, \ldots, n-1) \\ &\quad = \frac{\Pr(\text{risultati dei tentativi } 1, \ldots, n-1) \cdot p}{\Pr(\text{risultati dei tentativi } 1, \ldots, n-1)} \\ &\quad = p \end{aligned}$$

Dunque la probabilità di un successo al prossimo tentativo non dipende dai risultati visti finora. Cioè la teoria dei ritardi non serve a niente, ma avendo supposto inizialmente che le prove fossero indipendenti questo in realtà lo sapevamo già. Comunque, varie volte in passato ho trovato che questi calcoli convincono alcune persone; tipicamente, diciamo, 17-enni che non hanno pensato a ciò che stanno dicendo. Facendo questi calcoli superflui si convincono, e dicono "Argh! Che figura di merda che ho fatto!!"

Qualcuno sembra sostenere che due cose (indipendenti ecc. ecc.) che hanno sempre la stessa probabilità possano avere tempi medi di attesa diversi. Questo è davvero assurdo. Comunque … $p$ è sempre la probabilità di successo, $q = 1 - p$ quella d'insuccesso. Diciamo che abbiamo già fatto $n$ prove, e ne conosciamo i risultati. Mediamente, quanto dobbiamo aspettare per vedere il prossimo successo?

$$\begin{aligned} &P(\text{prossimo successo al turno } n+k | \text{risultati nei turni } 1, \ldots, n) \\ &\quad = \frac{P(\text{risultati nei turni } 1, ..., n \text{ poi } k-1 \text{ insuccessi poi un successo})}{P(\text{risultati nei turni } 1, \ldots, n)} \\ &\quad = q^{k-1} p \end{aligned}$$

Dunque il tempo medio che dovremo aspettare per il prossimo successo è

$$\begin{aligned} \sum_{k=1}^{\infty} k q^{k-1} p &= p + 2pq + 3pq^2 + \cdots \\ &= p\,(1 + q + q^2 + \cdots) + pq\,(1 + q + q^2 + \cdots) + pq^2\,(1 + q + q^2 + \cdots) + \cdots \\ &= p\,(1 + q + q^2 + \cdots)^2 = p/(1-q)^2 = 1/p. \end{aligned}$$

Vedete che il tempo medio di attesa del prossimo successo è sempre $1/p$. Una conoscenza dei risultati passati non cambia niente. I ritardi, essendo derivati dai risultati passati, non ci servono.

Vi sembrerà incredibile, ma c'è perfino chi dice "Ah, ma *io* scommetto sul fatto che il numero ritardato esca entro altre $m$ estrazioni." per qualche valore di $m$, dimostrando di non aver ancora capito che due copie dello stesso elenco di numeri non-negativi hanno la stessa somma. Vogliamo sprecare degli elettroni/toner/inchiostro? Va bene:

$$\begin{aligned} &P(\text{prossimo successo al turno } n+k | \text{risultati nei turni } 1, \ldots, n) \\ &\quad = \frac{P(\text{risultati nei turni } 1, ..., n \text{ poi } k-1 \text{ insuccessi poi un successo})}{P(\text{risultati nei turni } 1, \ldots, n)} \\ &\quad = q^{k-1} p \end{aligned}$$

Dunque la probabilità che il primo successo sia nei prossimi $m$ lanci è:

$$\sum_{k=1}^{m} q^{k-1} p.$$

Non figurano qui né $n$, né alcuna informazione sui primi $n$ risultati; quindi la probabilità di vincere nelle prossime $m$ estrazioni è sempre la stessa. Non importa quante estrazioni siano state fatte nel passato, o quali siano stati i loro risultati. Puoi sempre far finta che le prime $n$ estrazioni non siano state fatte.

## 2.2 La Legge dei Grandi Numeri

Una versione della legge dei Grandi Numeri dice, in linea di massima:
Supponiamo che un certo avvenimento si presenti con probabilità $p$ ad ogni prova (e che il risultato di ogni prova sia indipendente dai risultati di eventuali altre prove). Facciamo $n$ prove, e chiamiamo $S_n$ il numero di successi. Allora, con probabilità 1,

$$\lim_{n\to\infty} \frac{S_n}{n} = p$$

Se volete un'enunciato più preciso di questa e altre versioni della Legge dei Grandi Numeri vedete, per esempio, *Probability and Random Processes* di G.R. Grimmett e D.R. Stirzaker. Notate che ci sono vari sensi in cui una variabile casuale può convergere a qualcosa, e vari modi per paragonare due variabili casuali. Ma sono distinzioni che mi sembrano poco pertinenti qui.

La legge dei grandi numeri dice che la proporzione di successi tende (c.p. 1) a $p$. Ma spesso i ritardisti sembrano pensare che dica un'altra cosa, cioè che $S_n - np$ tende a zero. Non è la stessa cosa. Vediamo:

Consideriamo $x_n = n + \sqrt{n} + 30$ e $y_n = n$. $x_n/y_n$ tende a 1, ma $x_n - y_n$ tende a $\infty$.

Consideriamo anche $x_n = 2/n$ e $y_n = 1/n$. Qui $x_n - y_n$ tende a 0, ma $x_n/y_n$ tende a 2. Questo non è particolarmente pertinente, ma serve per mostrare che l'implicazione non vale in nessuno dei due sensi.

La legge dei grandi numeri non dice quello che i ritardisti vogliono che dica.

Anzi, il mio primo controesempio non è un caso patologico. Diciamo che abbiamo delle monete che escono testa con probabilità $p$. Sia $q = 1 - p$. Dopo $n$ lanci, il *numero* di teste avrà una distribuzione binomiale con media $np$ e standard deviation $\sqrt{npq}$. La *proporzione* di teste avrà media $p$ e standard deviation $\sqrt{pq/n}$. La proporzione di teste tende con probabilità 1 a $p$, ma la standard deviation del *numero* di teste va a infinito; quindi dire che la differenza fra $S_n$ e $p$ tende a 0 è sbagliato. Anzi, la probabilità che il numero di teste sia "vicino" ($|S_n - np| \leq d$, per qualche $d$, diciamo) a $np$ tende a zero. Se prendiamo $p = 1/18$ e diciamo che l'evento sia "esce un 6 a Roma" vediamo che questo è pertinente.

Alcuni ritardisti sofisticati dicono che se parti con troppa o troppo poca fortuna, allora ci sarà un periodo di bilanciamento, perché lo dice la L.D.G.N. Non funziona così. Basta che ci sia un periodo di normalità: la probabilità tende (con probabilità 1) a $p$ nel lunghissimo termine non perché avvenga necessariamente un "bilanciamento" ma solamente perché i risultati inziali vengono sommersi dai risultati normali seguenti. Se esce testa 4000 volte di fila e poi escono teste e croci alternanti all'infinito, cosa succede? La proporzione di teste tende a 1/2 senza alcun "bilanciamento" nel senso di un periodo in cui le teste escono meno spesso del normale.

Non siete convinti? Vi porto in una stanza con 65536 scatole. In ogni scatola ci sono 16 monete. Tutte le possibili successioni di testa e croce sono presenti. Sul coperchio di ogni scatola ho scritto i risultati dei primi 8 lanci in quella scatola. Vi invito a scrivere su ogni scatola la vostra previsione del numero di teste presenti nel secondo gruppo di 8 lanci, cioé quelli ancora sconosciuti. *Io* scriverei "4" su ogni scatola. Un ritardista che abusi della legge dei grandi numeri dice più o meno: "Ci devono essere 8 teste in ogni scatola, perché lo dice la legge dei grandi numeri. Allora conto le teste sul coperchio, $t$, e scrivo $8 - t$." Tramite un programma, o una versione più in piccolo di questo esperimento (con 4 scatole, diciamo), vedrete che il ritardista che abusa della legge dei grandi numeri sbaglia più spesso di me.

Dovrei dire qualcosa qui sui tempi di attesa? È superfluo quanto il resto, naturalmente. Alcuni ritardisti dicono cose quasi incomprensibili del tipo "Tutto quello che dice Atkinson è irrilevante: non punto sulla prossima uscita ma sulle *prossime* uscite." L'unico senso che riesco a dare a ciò è che secondo questo ritardista, nonostante il fatto che le probabilità di uscita di due numeri sono sempre uguali, uno ha un minore tempo medio di attesa dell'altro. Vedete la prossima sezione.

Notate anche che i limiti non funzionano come vorrebbero i ritardisti. Puoi cambiare qualsiasi numero finito di termini in una successione senza cambiare il suo limite. 400 mila miliardi è un numero finito.

Ecco un programma che lancia 50 monete (bilanciate), e poi chiede a me e ad un ritardista di prevedere il numero di teste in un secondo gruppo di 50 lanci. Io dico sempre 25, visto che su 50 lanci, è il numero più probabile di teste. Il ritardista, usando la sua versione personalizzata della Legge dei Grandi Numeri, dice che ci saranno 50 teste in tutti e 100 i lanci, e allora la sua previsione per il secondo gruppo di 50 è 50-(teste uscite nei primi 50 lanci). Questo programma stampa il numero dei suoi successi come proprizione del mio.

```
while (1) {

$heads=0;
$flag=0;

foreach(1..50) {
$num[$_]=int(rand(2));
$heads+=$num[$_];
}

$adam_prediction=$heads+25;
$ritardista_prediction=50;

foreach(51..100) {
$num[$_]=int(rand(2));
$heads+=$num[$_];
}

$trials++;
if ($adam_prediction==$heads) {
$adam_success++;
$flag=1;
}
if ($ritardista_prediction==$heads) {
$ritardista_success++;
$flag=1;
}

if($adam_success>0) {
if($flag==1) {
$ritardista_advantage=$ritardista_success/$adam_success;
print $ritardista_advantage."\n";
$flag=0;
}
}
}
```

Cosa dovrebbe uscire da questo programma? Il ritardista ti direbbe "Il numero stampato sarà maggiore di 1, perché la legge dei grandi numeri, sebbene parla di limiti all'infinito, ha poteri magici nel breve termine. Allora la mia previsione sarà corretta più spesso di quella di Adam". Non capisco la

Legge dei Piccoli Numeri usata dai ritardisti abbastanza bene da poter precisare meglio la previsione del ritardista. “Un qualche numero maggiore di 1”, diciamo.

Secondo me cosa dovrebbe uscire? Ok.. io dovrei avere ragione con probabilità $\binom{50}{25}\frac{1}{2^{50}}$ Il ritardista avrà ragione quando ci sono 0 teste poi 50, o 1 poi 49, ecc. ecc. Cioè con probabilità

$$\sum_{i=0}^{50}\left(\binom{50}{i}\frac{1}{2^{50}}\right)^2$$

Usando l'approsimazione di Stirling vediamo che il rapporto fra questi valori (ritardista/Adam) viene circa $1/\sqrt{2}$, anche sostituendo 50 con un più generale $2n$ e 25 con $n$. E infatti dopo molte iterazioni il programma stampa circa 0,7. Allora: Atkinson 1, ritardisti 0.

### 2.3 Equiprobabilità? Strategie ottimali?

Non è necessario supporre che i numeri abbiano la stessa probabilità di uscita. Magari per qualche motivo (palline più pesanti?) alcuni numeri escono più spesso di altri. L'indipendenza può valere ugualmente, come vale per la mia moneta che esce testa con probabilità $p$. Se i numeri del lotto hanno la stessa probabilità, e le estrazioni sono indipendenti, allora non importa quali numeri scegliete: la probabilità di vincere sarà sempere la stessa. La strategia ottimale, nel senso di avere il valore atteso più alto, è “non giocare”; ma facciamo finta che tu sia obbligato a giocare. Nel lotto italiano, grazie a differenze fra i moltiplicatori ecc., ci saranno giocate (cioè ambata vs. ambo ecc.) con un valore atteso meno negativo di altre. Nella lotteria inglese (e nel SuperEnalotto italiano, mi dicono) c'è un montepremi che va diviso fra chi sceglie 6 numeri corretti. Se sei costretto a comprare un biglietto per la lotteria inglese, dovrai scegliere combinazioni poco popolari: non cambia la probabilità di vittoria, ma se dovessi vincere, mediamente vincerai una proporzione maggiore del montepremi. Perciò la cosa giusta da fare sarebbe forse di studiare tutte le teorie assurde, informarsi sui numeri più giocati, ecc. e poi generare delle combinazioni a caso, buttando quelle contenenti numeri troppo giocati (grazie a compleanni, ritardi, consigli dati in TV dai pirla, ecc.). Nota che la probabilità che una combinazione sia giocata non sarà uguale al prodotto delle probabilità che siano giocati i suoi componenti: potrà quindi capitare che alcune combinazioni rarissime contengano numeri popolari.

Se alcuni numeri sono più probabili di altri, ma le estrazioni sono indipendenti, allora tanto vale puntare sui numeri più probabili se sei obbligato a giocare. (Nel caso della moneta, se $p > 0.5$ conviene puntare su testa). In questo caso i ritardisti hanno anche più torto del solito — se ritieni che i numeri abbiano probabilità diverse, ti conviene (beh, perdi di meno) puntare su quelli più probabili — e i numeri con ritardi alti saranno solitamente quelli con probabilità basse.

In realtà, non pare che ci siano motivazioni per credere che le probabilità di uscita dei singoli numeri siano molto diverse.

### 2.4 Ritardo massimo teorico?

Alcuni ritardisti sembrano credere che ci sia un massimo ritardo teorico. Messi sotto pressione, alcuni diranno che nonostante il nome non è un massimo teorico. Altri dicono, invece, che è davvero invalicabile. Boh. Comunque, almeno all'interno del modello che stiamo usando in questa sezione, non cè alcun massimo ritardo teorico (Beh, naturalmente, il ritardo massimo non può essere maggiore del numero di estrazioni). Anzi, prima o poi, uscirà, con probabilità 1, qualisiasi successione finita di risultati che vuoi. Si possono calcolare i tempi di attesa con le catene di Markov. Partendo da zero, il tempo di attesa per 220 insuccess di fila con un dato numero in una data città è circa 5.205.463,7 estrazioni. Visto che nel lotto italiano ci sono 900 numeri diversi, è chiaro che il massimo ritardo storico dovrebbe essere (mediamente) maggiore di quello ottenuto nello stesso numero di estrazioni guardando un numero solo. So calcolare la distribuzione del massimo ritardo con un solo numero dopo $n$ estrazioni, ma non so calcolarla per un'intera città (o tutta la lotteria italiana): i 90 numeri in una data città non vivono in 90 copie indipendenti dello stesso processo stocastico, visto che esattamente 5 dei numeri escono in ogni estrazione. Non so come trattare questo problema dal punto di vista teorico.

Quanto dobbiamo aspettare per il primo ritardo di lunghezza $r$?

$$media = \frac{1-q^r}{pq^r} = \frac{1}{p}(\frac{1}{q^r} - 1)$$

(add explanation of how to get this - only normal Markov Chain stuff)

$$s.d. = \sqrt{\frac{1}{(pq^r)^2} - \frac{2r+1}{pq^r} - \frac{q}{p^2}}$$

(don't bother explaining how to get this? damned complicated)

Per $r$ grande, s.d. $\approx$ media.

Notate che i tempi di attesa di successioni miste si comportano in modo abbastanza strano: è possibile nominare due successioni S e T della stezza lunghezza tale che S ha un minore tempo di attesa di T ma se aspetti entrambi allo stesso tempo, probabilmente sarà T la prima ad apparire.

(add examples)

Notate però che tutto ciò può essere interessante, ma è inutile. Non ci permette di prevedere le singole estrazioni.

## 3 I ritardisti "dipendentisti"

La teoria dei ritardi può funzionare solamente se gli eventi in questione *non* sono indipendenti. Perciò un ritardista, se vuole essere preso sul serio, è obbligato a dire "Le estrazioni del lotto non sono indipendenti". Questa posizione è coerente, anche se quasi sicuramente difficile da difendere. Nella mia esperienza, pochissimi ritardisti lo dicono. Ma non si sa mai.

In questo caso tutti i miei calcoli e ragionamenti precedenti, basati sugli eventi indipendenti, non valgono più, perchè il modello usato non corrisponde alla presunta situazione reale.

Possiamo benissimo chiederci quando è valido ritenere che degli eventi siano indipendenti.

Se scelgo delle carte da un mazzo di carte e le brucio, sto facendo una serie di prove non-indipendenti. Se la prima carta è l'asso di cuori, so che le altre carte non possono mai essere l'asso di cuori. In questo caso le informazioni sul passato mi danno informazioni sul futuro. Se ho bruciato 20 carte rosse e 5 carte nere, e qualcuno mi chiede di prevedere il colore della prossima carta, dico "nero".

Se qualcuno mi chiede di prevedere la temperatura domani, è utile sapere che tempo fa oggi: anche qui l'indipendenza non c'è.

Ma se qualcuno lancia due dadi diversi in stanze diverse, ritenere che il numero su un dado influenzi quello sull'altro sarebbe molto strano. E un solo dado o una sola moneta lanciata più volte? Come "sa" cosa è successo nel passato? Un mazzo di carte semidistrutto ha una certa memoria: alcune carte non ci sono più. Ma le monete non hanno memoria.

Sembra plausibile che ci sia molta dipendenza nelle estrazioni del lotto? È difficile che le macchine del lotto si influenzino a distanza; sembra quindi plausibile dire che i risultati a Roma e quelli a Napoli sono indipendenti. (Se credi che le estrazioni in posti diversi siano dipendenti, ti chiedo: come sa una macchina del lotto che deve farsi influenzare solamente da altre macchine del lotto in Italia? E che effetto hanno le macchine del lotto non attualmente in servizio? Potresti, forse, costruire una macchina del lotto a casa tua e farla funzionare molte volte allo scopo di influenzare le altre macchine del lotto in Italia?)

Forse le estrazioni nello stesso posto sono in qualche modo dipendenti? Questo non è totalmente implausibile: se il meccanismo è una ruota molto arrugginita, forse i numeri scelti si comportano come la temperatura in giorni successivi. Può darsi che la macchina non riesca a generare numeri molto diversi, perché gira con molta difficoltà. O magari nel caso del lotto italiano, i numeri usciti la settimana scorsa vengono messo in fondo all'urna, gli altri numeri vengono messi sopra, e poi le palline nell'urna non sono mischiate. Questo creerebbe della dipendenza.

Comunque, se credi nella dipendenza fra i risultati successivi di un singolo dado, di una singola moneta o di una singola macchina-lotto, hai altri problemi: cosa succede se qualcuno fa funzionare la macchina del lotto per verificarne il funzionamento? Rovina tutto!

Se gli operai nella zecca di stato lasciano cadere delle monete, sono già state lanciate! Oddio! Non puoi fare più i tuoi calcoli! Allora mi converrebbe fare così: prendo delle monete dalla zecca. Ne lancio una finché non esca testa dieci volte di fila. La metto in una scatola. Ripeto per ogni moneta. Poi vendo ai ritardisti queste monete pre-lanciate-con-ritardo-incorporato. Le comprerebbero? Tu le compreresti? Se credi nella dipendenza fra le estrazioni di una singola macchina del lotto, dovrebbe voler dire che se io prendo una macchina del lotto, la faccio andare qualche volta, e poi la passo a te, tu sarai in grado di dedurre qualcosa sui numeri già usciti guardando quelli che escono quando la usi tu. Ti sembra plausibile?

Un'altra possibilità: forse i contenuti di una singola estrazione di una singola ruota sono più complicati di quanto pensiamo. Magari la probabilità di una data combinazione di 5 numeri non è uguale al prodotto delle loro probabilità di uscita. Dopo tutto, quando facciamo un'estrazione, non stiamo necessariamente scegliendo 5 numeri da 90 senza rimpiazzo, con risultati indipendenti. Magari la dipendenza sta qui. Forse il 3 e il 6 non escono mai insieme (o qualcosa del genere). Dovremmo conoscere le probabilità di tutte le 43.949.268 combinazioni possibili e poi controllare se è possibile ricavarle usando solamente le probabilità dei 90 numeri. Questo lo dovremmo rifare per ogni singola ruota. E dovremmo rifarlo ogni volta che una ruota viene cambiata, aggiustata, ecc. Farlo sulla base dei risultati storici del lotto in Italia è impossibile, perché non sono state fatte centinaia di milioni di estrazioni. E se fossero state fatte, le macchine si sarebbero guastate e avremmo dovuto ricominciare.

Forse è opportuno segnalare qui il paradosso dei compleanni. (La probabilità che fra 23 persone ci siano almeno 2 con lo stesso compleanno è maggiore di 0,5). Dopo quante estrazioni c'è una probabilità maggiore di 0,5 di aver avuto almeno una cinquina ripetuta? Usando lo stesso ragionamento che si usa per i compleanno, si può vedere che la risposta è 7806. Sorprendentemente basso.

È chiaro che l'indipendenza all'interno della singola estrazione, l'equiprobabilità, l'indipendenza fra le diverse ruote e fra successive estrazioni su una sola ruota fanno parte di un modello che cerchiamo di fare della realtà. Dobbiamo chiederci se questo modello corrisponde o no alla realtà. Secondo me nel caso del lotto un modello così non dovrebbe essere troppo diverso dalla realtà. E molti ritardisti sarebbero d'accordo.

Il problema è che il lotto è un gioco così svantaggioso che ben difficilmente gli effetti di ipotetiche piccolissime differenze di peso fra le palline, o la dipendenza nei risultati in una singola città, potranno farlo diventare un gioco favorevole. Poi, se il lotto fosse favorevole per il giocatore, non esisterebbe: lo Stato ha creato il lotto per guadagnare, non per perdere. Nello stesso modo, se i giochi di casinò fossero favorevoli, non esisterebbero. (Nel caso del blackjack, in cui alcune versioni ideali *sono* favorevoli, succede che nei casinò le regole del gioco sono state cambiate e/o chi cerca di giocarlo bene viene buttato fuori. Non sono mai stato in un casinò, quindi non so come si gioca a blackjack nel mondo reale.)

## 3.1 Il ritardo invalicabile

È irragionevole ritenere che ci sia un ritardo massimo invalicabile. Non mi è chiaro che ci sono persone che ritengono questo o no. Comunque, analizziamo in modo finto serioso un modello del lotto in cui c'é un ritardo invalicabile. Chiaramente, questa discussione *d*eve stare nella sezione "dipendentisti".

Il numero non importa. Chiamiamolo $m$, per "massimo".

Allora quello che un ritardista vuole sapere su una data routa sono i ritardi dei suoi 90 numeri. Non importa altro. Allora per specificare lo stato della routa, servono 90 interi fra 0 e $m$. Questi sono i ritardi di 1, 2, 3, ... .

Dopo ogni unita' di tempo, èfatta un'estrazione, e 5 numeri escono. I ritardi dei numeri appena usciti diventano 0 (o rimangono 0) e gli altri 85 aumentano di 1.

L'impossibilita' di avere un ritardo di $m+1$ rende impossibili anche molti altri stati. Per esempio, non possiamo avere piùdi 5 numeri con un ritardo di $m$, piùdi 10 con ritardo $m-1$ o superiore, più' di 15 con ritardo $m-2$ o superiore, ecc. ecc. ... , o più di 85 con ritardo $m-16$ o superiore. (per avere 85 numeri con ritardo $m-16$, basta che gli stessi 5 numeri escono $m-16$ volte di fila)

Se all'ipotetico ritardista non piace l'idea che la stessa cinquina esca $m-16$ volte di fila, forse vorra' vietare anche altri stati. Nessun problema. Potremmo far alternare due cinquine, o qualcosa del genere.

Allora, con questo modello, è possibile costruire una situazione in cui il modello $g$arantisce le prossime 16 o 17 cinquine. Questo è ovviamente balordo.

# 4 Doppino, progressioni, ecc.

Non funzionano. Non dare retta a chi dice diversamente. Se usi uno di questi "sistemi" per n turni, la tua vincita media sarà quella che sarebbe stata con qualsiasi altra strategia. Naturalmente, se smetti di giocare dopo l'$n$-esima vittoria, avrai vinto qualcosa, ma a meno che tu non abbia capitali infiniti, non puoi garantire di poter giocare fino all'$n$-esima vittoria.

(Altra roba da aggiungere qui)

# 5 Tempi di attesa

Ritardista: I numeri in ritardo devono uscire prima o poi

Matematico: Tutti i numeri devono uscire prima o poi. (beh, "devono" nel senso di "con probabilità 1")

Ritardista: Ma se qualcosa deve succedere prima o poi, e la aspetti da un po', allora vuol dire che adesso hai meno tempo da aspettare, no? Allora i numeri in ritardo usciranno prima di altri numeri.

Matematico: Non necessariamente ...

Ovviamente, per il lotto ho già risposto a questo nella sezione "Indipendentisti". Comunque, vediamo cosa può succedere in generale coi tempi di attesa. Qui userò un sacco di formule, perché non vedo alternative. Se non capisci l'algebra, imparala (non è così difficile) o salta questa sezione.

Per ora, consideriamo cose che "devono succedere prima o poi". Consideriamo una serie di giochi così: io ho delle carte numerate (1,2,3, ... ), e una di queste carte è quella vincente (c'è un simbolo speciale sopra, o qualcosa del genere). Tu giri le carte finché non vinci. Come scelgo la carta vincente? Genero un intero positivo a caso. Ma "a caso" potrebbe voler dire più o meno qualsiasi cosa. Faccio in modo che la carta vincente sia $i$ con probabilità $f(i)$. Visto che (per ora) devi vincere prima o poi, sappiamo $\sum_{i=1}^{\infty} f(i)=1$.

Per ora, diciamo anche che il tempo medio di attesa
'e finito. Così potremo vedere se sta aumentando, diminuendo, o cosa.

Allora

$$\sum_{i=1}^{\infty} ip(i)<\infty$$

## 5.1 Tempi di attesa che diminuiscono

La carta vincente è una fra 1,2, ... ,$a$, tutte con probabilità $1/a$. Il tempo medio di attesa è inizialmente $(a+1)/2$.

Diciamo che tu hai girato $n$ carte, e non hai ancora vinto. Mediamente, quante altre carte dovrai girare?

$$\sum_{i=1}^{\infty} i \frac{\operatorname{Pr}(\text{ primo successo dopo } n \text{ a } n+i)}{\operatorname{Pr}(n \text{ insuccessi iniziali})}$$

E questo viene

$$\sum_{i=1}^{a-n} i \frac{1}{a-n} = \frac{a-n+1}{2}$$

allora il tempo medio che dovrai aspettare ancora diminuisce

### 5.2 Tempi di attesa che non cambiano

La carta vincente è $i$ con probabilità $pq^{i-1}$, dove $p$ e'un qualche numero fra 0 e 1, e $q = 1 - p$. Questa è la distribuzione geometrica, e abbiamo già visto che il tempo medio di attesa è costante.

### 5.3 Tempi di attesa che crescono

La carta vincente è $i$ con probabilità $\frac{1}{2}(p_1 q_1^{i-1} + p_2 q_2^{i-1})$

La probabilità di non vincere entro il giorno $n$ è $\frac{1}{2}(q_1^n + q_2^n)$

La probabilità di vincere dopo altre $k$ carte se ne hai girato $n$ senza successo è $\Pr(\text{successo a } n + k)/\Pr(\text{insuccessi fino a } n)$

Dunque il tempo medio che dovrai aspettare ancora è:

$$\sum_{k=1}^{\infty} k \frac{\frac{1}{2}(p_1 q_1^{n+k-1} + p_2 q_2^{n+k-1})}{\frac{1}{2}(q_1^n + q_2^n)}$$
$$= \frac{1}{q_1^n + q_2^n}\left(\frac{p_1 q^n}{(1-q_1)^2} + \frac{p_2 q_2^n}{(1-q_2)^2}\right)$$
$$= \frac{q_1^n \frac{1}{p_1} + q_2^n \frac{1}{p_2}}{q_1^n + q_2^n}$$

E questo cresce col crescere di $n$, tendendo al maggiore di $1/p_1$ e $1/p_2$

Un esempio nel quale i tempi di attesa tendono a $\infty$:

Cominci ad aspettare a tempo $t = 1$. Il premio appare ad un qualche tempo futuro. La densità è $\frac{n-1}{t^n}$, per un qualche $n > 2$.

La probabilità di non aver vinto entro $t = r$ è $\int_r^\infty \frac{n-1}{t^n} dt = \frac{1}{r^{n-1}}$

Dunque, il tempo medio che dovrai aspettare ancora, dato che hai gia' aspettato fino a $r$ è:

(hmm. orribile)

### 5.4 Tempi di attesa che fanno un po' di tutto

Any old thing

### 5.5 Tempi di attesa infiniti, cose che possono non succedere mai, ecc.

Prendiamo una moneta che esce "testa" con probabilità 0,5. Ovviamente, inizialmente il numero di teste è uguale al numero di croci: sono entrambi 0. Che probabilità c'è che i numeri si ugualino prima o poi una volta che cominciamo a lanciare la moneta? 1. Dopo quanto tempo, mediamente, succederà? La risposta sorprendente è "infinito".

random walk in 1, 2, 3 dimensions

## 6 Attendibilità

Alcuni ritardisti potrebbero cominciare a dire che loro non parlano di probabilità, ma di "attendibilità". L'attendibilità e' una cosa fittizia inventata dal santone/guru dei ritardisti, Leontino Gorgia.

Lui definisce l'attendibilità tramite una formula il cui significato non é affatto chiaro.

$$A = \frac{1}{1+90q^r}$$

dove $q$ è 17/18, $r$ il ritardo, e $A$ l'attendibilità. Il "senso" di questa formula sarebbe: $90q^r$ è la quantità di numeri su una ruota che aspettiamo di trovare ad un ritardo di almeno $r$. Immagino che nel caso dell'intera lotteria italiana bisogni mettere 900 al posto di 90.

Dice (correttamente) che i valori che escono dalla formula saranno sempre fra 0 e 1. Ma questi valori si chiamano "attendibilità" solamente perché a lui piaceva quel nome. Avrebbe potuto chiamare il valore "vombatitudine" o "grado di somiglianza ad un cruscotto". Il fatto che lui ha scelto il nome "attendibilità" non vuol dire che il calcolo abbia alcuna utilità o significato.

Inventare una formula balorda che dia valori fra 0 e 1 non è molto difficile. Inventare una formula balorda che manda ritardi alti ad "attendibilità" alte non è molto difficile. Perché questa formula invece di un'altra? Beh, perché non deve avere alcun significato, sta lì solamente per impressionare la gente numericamente analfabeta.

Ritardista di Turno #2 dice che le probabilità non cambiano, e che l'attendibilità sia solamente un criterio per scegliere. Questo è strano ma ok: RDT2 sceglie dei numeri non perché abbiano una maggiore possibilità di vincere, ma in base ad un altro criterio, cioé una formula balorda. Potrebbe scegliere il compleanno della madre, o la temperatura ad Oslo, ma sceglie il numero con maggiore "attendibilità". Visto che tutti i numeri hanno la stessa probabilità di uscire, questo metodo vale quanto qualsiasi altro metodo (cioè zero), ma sembra inutilmente complicato.

# 7 Mettere il ritardismo alla prova

## 7.1 $n$ numeri equiprobabili

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <time.h>

int n=90;
int i,newnum,maxdelay,choose,trials,success;
double advantage;
int seen=0;
int delay[91],happened[91];

int main(void) {

long t;
double sdbase=sqrt((double)(n-1)/n/n);
double sd;
double z;
srand48(time(&t));

printf("Un indice di successo di k vuol dire che il numero piu' in ritardo e' \
uscito con una frequenza k volte quella di un numero qualsiasi. In questo programma,\
il numero qualsiasi e' 1\n\n");

printf("Secondo il mio modello, Z dovrebbe essere fra -1,96 e 1,96 \
con probabilit 95%\n");
```

```
for (i=1;i<=n;i++) {
 delay[i]=0;
 happened[i]=0;
 }

while (seen<n) {
 newnum=1+(int)floor(drand48()*n);
 delay[newnum]=0;
 if (happened[newnum]==0) {
   happened[newnum]=1;
   seen++;
   }
 for (i=1;i<=n;i++) {
  if (i!=newnum) {
   delay[i]++;
   }
 }

 }

while (trials<=10000000) {

maxdelay=0;
choose=0;
for (i=1;i<=n;i++) {
 if (delay[i]>maxdelay) {
  choose=i;
  maxdelay=delay[i];
  }
 }

 newnum=1+(int)floor(drand48()*n);
 delay[newnum]=0;

for (i=1;i<=n;i++) {
 if (i!=newnum) {
  delay[i]++;
  }
 }

trials++;
if (choose==newnum) {
 success++;
 }

if (trials%100000==0) {
 advantage=(double)success/trials*n;
 sd=n*sdbase/sqrt((double)trials);
 z=(advantage-1)/sd;
 printf("Iterazioni: %i  Indice di successo dei ritardi: %g  Z: %g\n",\
trials, advantage, z);
 }
}
```

```
}
```

### 7.2 $n$ numeri con probabilità diverse

```
$n=10;

$common=0;

$top[0]=0;
$top[$n]=1;
$min_p=1;

for(1..$n-1) {
$top[$_]=rand;
}

@top = sort @top;

for(1..$n) {
$p[$_]=$top[$_]-$top[$_-1];
if ($p[$_]>$max_p) {
$max_p=$p[$_];
        $best=$_;
        }
    if ($p[$_]<$min_p) {
     $min_p=$p[$_];
        $worst=$_;
        }
}

while(1) {

if ($common==0) {
$adam=1;
    $oldest=1;
    $last=1;
    } else {
    $adam=$common;
    $ritardista=$oldest;
    $another=$last;
    }
$dio=$best;
$idiot=$worst;

$x=rand;
$i=1;

while($x>$top[$i]) {
$i++;
    }

$trials++;
```

```
if ($i==$adam) {
$adam_s++;
    }

if ($i==$ritardista) {
$rita_s++;
    }

if ($i==$dio) {
$dio_s++;
    }

if ($i==$another) {
$anot_s++;
    }

if ($i==$idiot) {
$idio_s++;
    }

if ($trials%100==0) {
print $dio_s/$trials." ".$adam_s/$trials." ".$anot_s/$trials." ";
print $rita_s/$trials." ".$idio_s/$trials."\n";
}

$last=$i;

$count[$i]++;
if ($count[$i]>$count[$common]) {
$common=$i;
print "Adam changes at trial ".$trials."\n";
}

$max_age=0;

for(1..$n) {
    if($_==$i) {
     $age[$_]=0;
     } else {
     $age[$_]++;
     if($age[$_]>$max_age) {
     $max_age=$age[$_];
         $oldest=$_;
         }
     }
}

}
```

### 7.3 La legge impotente dei piccoli numeri

## 8 Case studies

### 8.1 Samaritani

Non sembra un ritardista. Da quello che dice nel suo libro (trovabile nella Biblioteca Nazionale) sembra uno che ha fatto un po'di grossolanometria: quello che ha detto è stato interpretato male dai ritardisti. Non *c*redo che Samaritani avrebbe detto che gli eventi indipendenti non fossero indipendenti.

### 8.2 Ritardisti di Turno #1-4

## 9 Analfabetismo matematico puro e semplice

Numeri adiacenti ecc.

Ci sono stati più incidenti con gli aerei nel 1997 che nel 1910, ma questo non vuol dire che gli aerei moderni sono meno sicuri.

Se sparo ad una folla, e una persona su cento che muore è bionda, non vuol dire che i biondi sono a prova di proiettile. Magari solamente una persona su cento era bionda.

## 10 I ritardisti sono sinceri?

### 10.1 Ignoranza

non sapere qualcosa. diagnosi implausibile

### 10.2 Stupidità

non sapere una cosa, e essere incapaci di impararla. alcuni ritardisti fingono di essere stupidi, ma non credo che molti lo siano.

### 10.3 Baker Street Irregulars e/o finti tonti

sapere perfettamente che una cosa è falsa, ma fare finta di crederci. Per esempio i *Baker Street Irregulars* fanno finta di credere che Sherlock Holmes sia esistita, ma è tutto un gioco. Non ci credono davvero. I soci della *Flat Earth Society* decenni fa erano sinceri, ma adesso è uno scherzo - scrivono articoli finti per gioco, cercando di spiegare ciò che si vede nel cielo partendo dal presupposto che il mondo sia piatto.

Qualcuno sostiene che i ritardisti siano come i Baker Street Irregulars. Questo mi sembra implausibile. Comunque, la diagnosi più plausible, imho, è che fanno i finti tonti. Sanno perfettamente che raccontano balle. L'attendibilità, il namedropping della Legge dei Grandi Numeri, il Teorema del Limite Centrale, il Teorema di Gödel, ecc. sono tutti specchietti per le allodole. Come si dice in italiano "blinding with science"? Comunque, devono raccontare delle balle pseudo-matematiche che convinceranno le loro vittime matematicamente analfabete a comprare programmi/riviste ecc.

## References

[1] W. Feller, An Introduction to Probability Theory and its Applications, 1968

[2] G. R. Grimmett, D.R. Stirzaker, Probability and Random Processes, 1992

[3] Paulos, Innumeracy